KOOC W3

Continuous Optimization (II) constrained

EE, KAIST

김창익

Categorization

- Categorization of continuous optimization
  - Unconstrained optimization
    - ~~Linear~~
    - Nonlinear
  - Constrained optimization
    - Linear programming
    - Nonlinear programming

Categorization

- Unconstrained optimization:

$$\boldsymbol{x}^* = \min f(\boldsymbol{x})$$

$$subject\ to\ \boldsymbol{x} \in \Omega,$$

$$\boxed{\text{where } \Omega = \boldsymbol{R}^n.}$$

- Constrained optimization:
  - feasible set is the form of functional constraints.

$$\Omega = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = 0, \boldsymbol{g}(\boldsymbol{x}) \le 0\}.$$

Contents

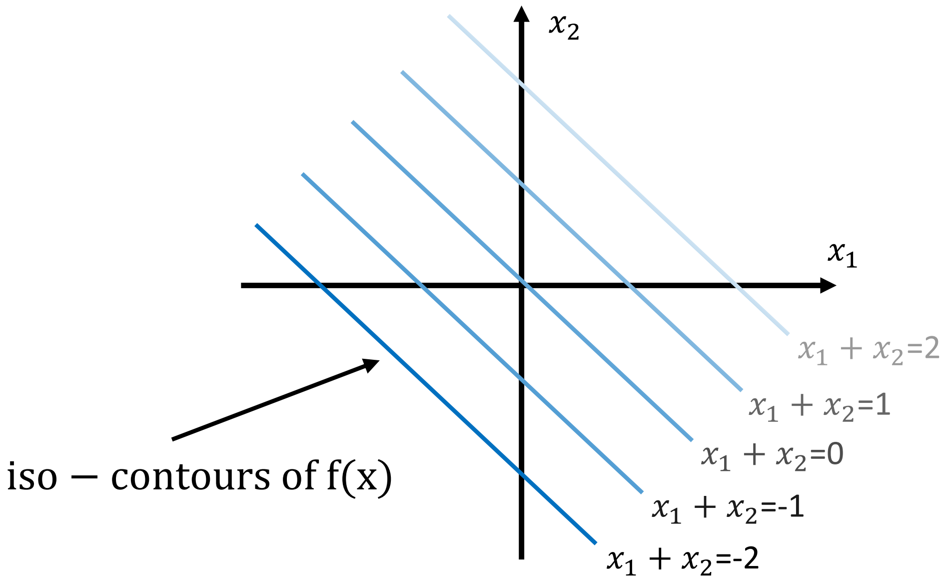
Contents

- Linear programming
- Nonlinear programming
  - Lagrange multipliers methods
    - Method of Lagrange multiplier: equality constraints
    - Lagrangian Dual Problem
    - KKT Conditions
  - Projection-type methods
  - Penalty-type methods

Linear programming

- Linear programming (1940s):
  - The objective function and all of the constraints are affine.
  - That is, $a^T \mathbf{x} + b, a \in R^n, b \in R$
  - For reference, linear functions are in the form of $a^T \mathbf{x}$. i.e., $b = 0$

$$
\begin{array}{ll}
\text{maximize} & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\
\text{subject to} & A\mathbf{x} \leq \mathbf{b} \\
\text{and} & \mathbf{x} \geq \mathbf{0}
\end{array}
$$



$$f(\mathbf{x}) = x_1 + x_2$$

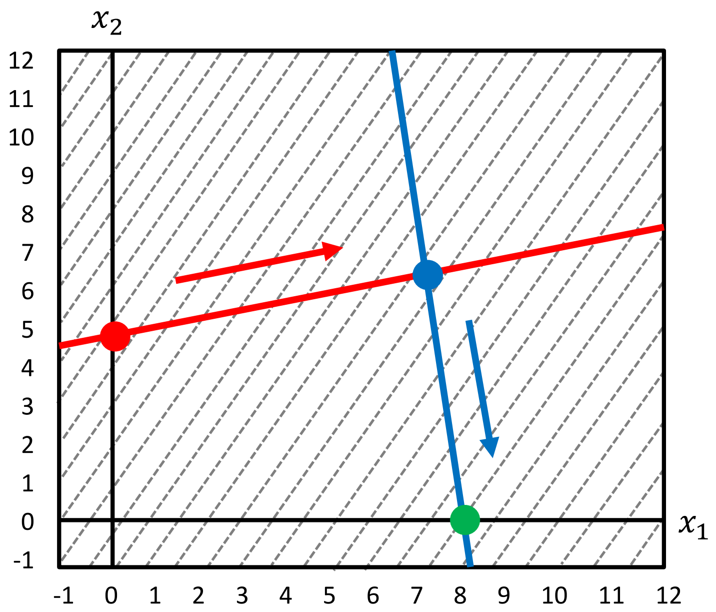
Linear programming

- Linear programming (1940s):

Ex) minimize $f(x) = 3 \cdot x_1 - 1.5 \cdot x_2$

subject to $-0.3 \cdot x_1 + x_2 \leq 5$

$6.25 \cdot x_1 + x_2 \leq 50$



Methods for Constrained Optimization

- A certain class of algorithms for solving constrained optimization problems.
  - Penalty method: replaces a constrained optimization problem by a series of unconstrained problems whose solutions ideally converge to the solution of the original constrained problem. Penalizing any violation of the constraints.

$$\boxed{\begin{aligned} &\min_{\boldsymbol{x}\in R^n} f(\boldsymbol{x}) \\ &subject\, to\; h_i(\boldsymbol{x}) = 0,\; i = 1,\ldots, m \\ &\qquad\qquad\quad g_i(\boldsymbol{x}) \le 0,\; j = 1,\ldots p. \end{aligned}}$$

$$\Rightarrow \min_{\boldsymbol{x}\in R^n} q_k(\boldsymbol{x}) = f(\boldsymbol{x}) + c_k\left(\sum_{i=1}^{m} h(\boldsymbol{x})^2 + \sum_{j=1}^{p} \max(0, g_j(\boldsymbol{x}))^2\right)$$

  - Barrier method (Interior point method): a barrier penalty term that prevents the points generated from leaving the feasible region is added to the objective function.
  - Lagrange multipliers methods: next slides

Method of Lagrange Multiplier

- The method of Lagrange multipliers is a strategy for finding the local maxima and minima of a function subject to equality constraint.
- Consider an optimization problem

$$\textit{maximize}\ f(x)$$
$$\textit{subject to}\ h(x) = 0.$$

- We need both f and h to have continuous first partial derivatives.
- Define the Lagrangian as $\mathcal{L}(\mathbf{x}, \mu) = f(\mathbf{x}) + \mu\, h(\mathbf{x})$

Then $\mathbf{x}^*$ a local minimum $\Longleftrightarrow$ there exists a unique μ^* s.t.

1. $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \mu^*) = \mathbf{0}$ → $-\nabla_x f(x^*) = \mu^* \nabla_x h(x^*)$
2. $\nabla_{\mu} \mathcal{L}(\mathbf{x}^*, \mu^*) = 0$ → the equality constraint $h(x^*) = 0$
3. $\mathbf{y}^t (\nabla^2_{\mathbf{xx}} \mathcal{L}(\mathbf{x}^*, \mu^*))\mathbf{y} \geq 0 \quad \forall \mathbf{y}$ s.t. $\nabla_{\mathbf{x}} h(\mathbf{x}^*)^t \mathbf{y} = 0$

↑

Positive definite Hessian tells us we have a local minimum

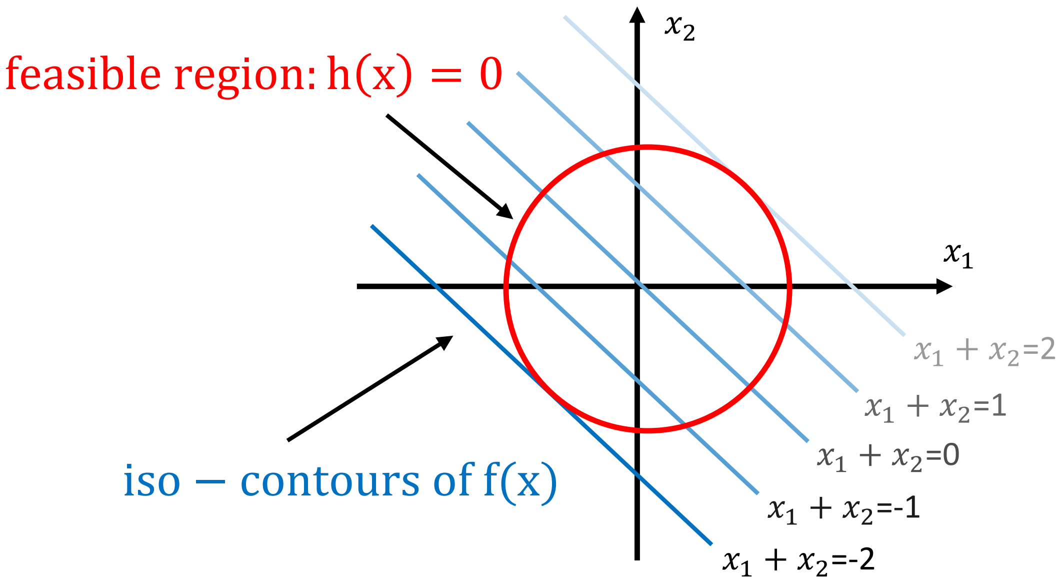
Method of Lagrange Multiplier

- Example (1/6)

$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \text{ subject to } h(\mathbf{x}) = 0}$$

where

$$\boxed{f(\mathbf{x}) = x_1 + x_2 \text{ and } h(\mathbf{x}) = x_1^2 + x_2^2 - 2}$$

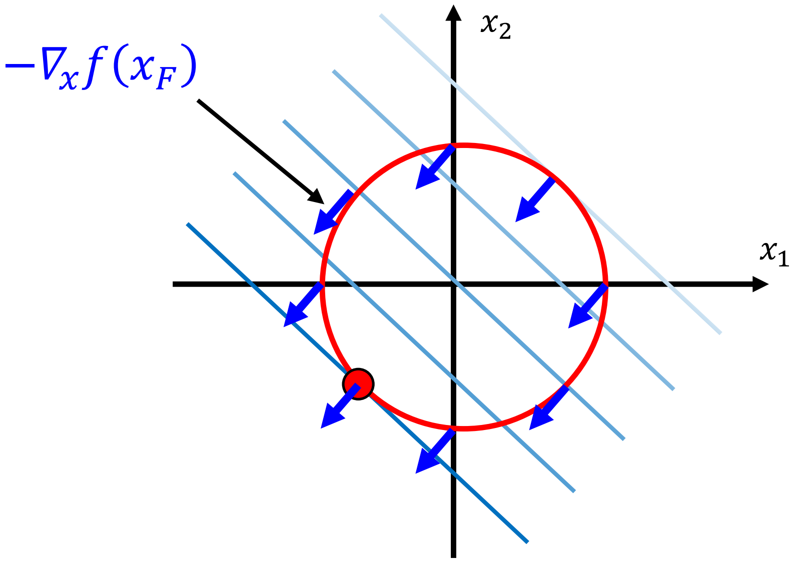


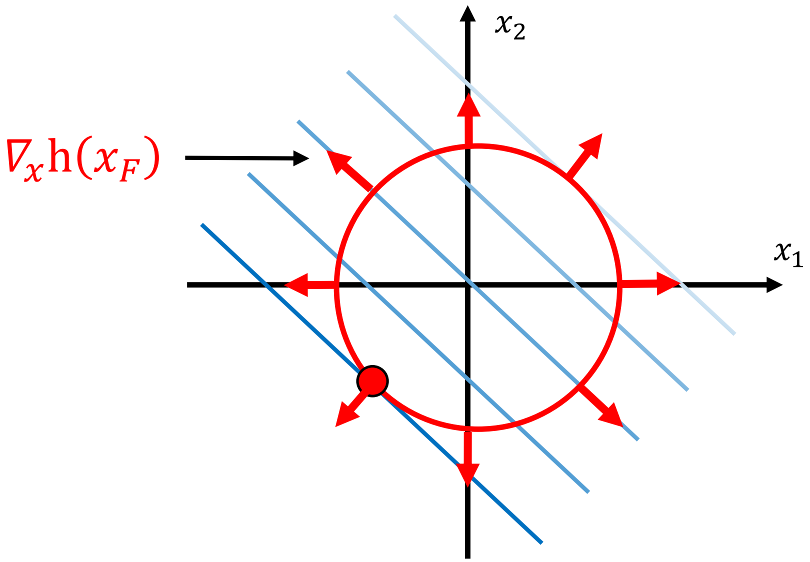
Method of Lagrange Multiplier

- Example (2/6)

Lagrangian $L(\mathbf{x},\mu) = f(\mathbf{x}) + \mu h(\mathbf{x}) = x_1 + x_2 + \mu(x_1^2 + x_2^2 - 2)$

$$-\nabla_{\mathbf{x}} f(\mathbf{x}^*) = \mu^* \nabla_{\mathbf{x}} h(\mathbf{x}^*)$$

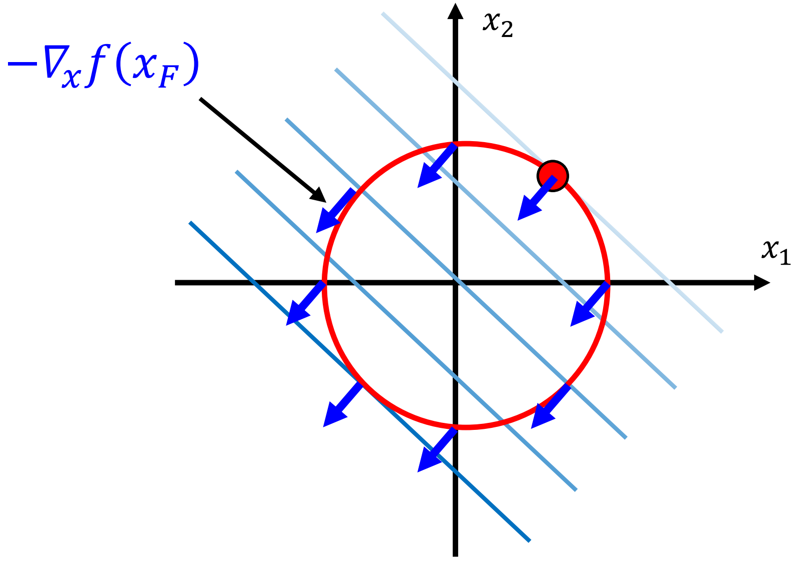


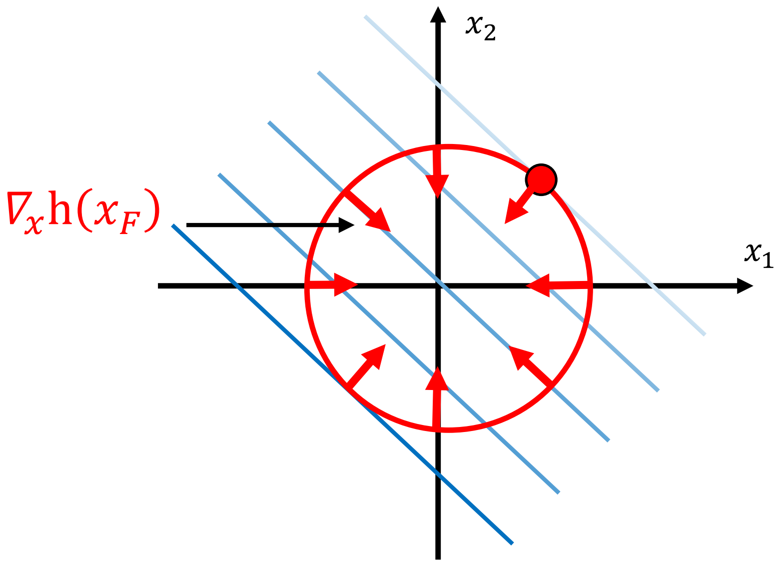


Method of Lagrange Multiplier

- Example (3/6)

$$-\nabla_x f(\mathbf{x}^*) = \mu^* \nabla_x h(\mathbf{x}^*)$$

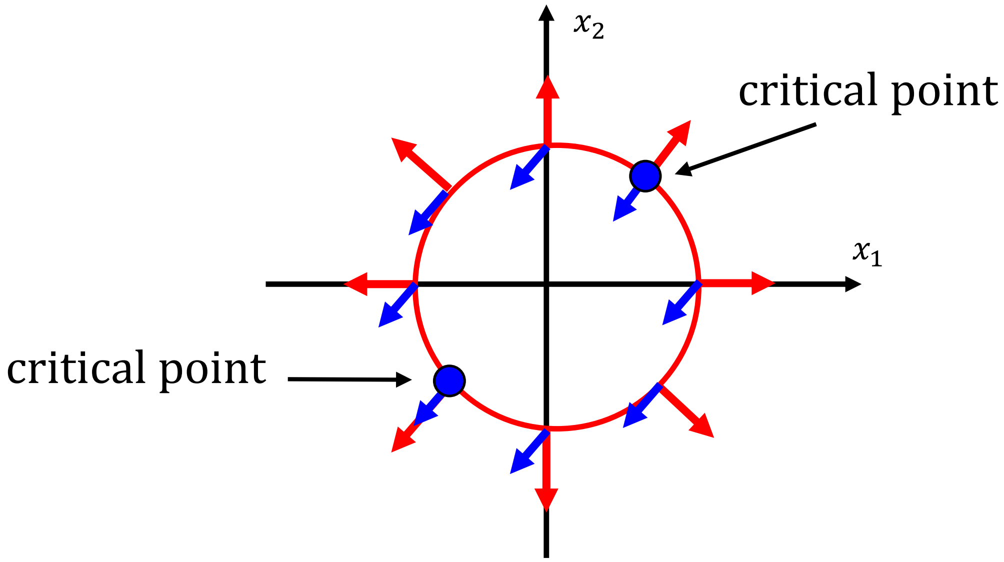




Note the direction of the normal is arbitrary as the constraint be imposed as either $h(x) = 0$ or $-h(x) = 0$

Method of Lagrange Multiplier

- Example (4/6)



A constrained local optimum occurs at x^* when $\nabla_x f(x^*)$ and $\nabla_x h(x^*)$ are parallel that is

$$-\nabla_x f(\mathbf{x}^*) = \mu^* \nabla_x h(\mathbf{x}^*)$$

Method of Lagrange Multiplier

- Example (5/6)

$$\frac{\partial L}{\partial x_1} = 1 + 2\mu x_1 = 0, \quad \rightarrow x_1 = -\frac{1}{2\mu}$$

$$\frac{\partial L}{\partial x_2} = 1 + 2\mu x_2 = 0, \quad \rightarrow x_2 = -\frac{1}{2\mu}$$

$$\frac{\partial L}{\partial \mu} = x_1^2 + x_2^2 - 2 = 0$$

$$\frac{1}{4\mu^2} + \frac{1}{4\mu^2} = 2$$

$$\therefore \mu^* = \pm\frac{1}{2}.$$

$$For\ \mu = \frac{1}{2}, \quad x_1 = x_2 = -1,$$

$$For\ \mu = -\frac{1}{2}, \quad x_1 = x_2 = 1.$$

Method of Lagrange Multiplier

- Example (6/6)

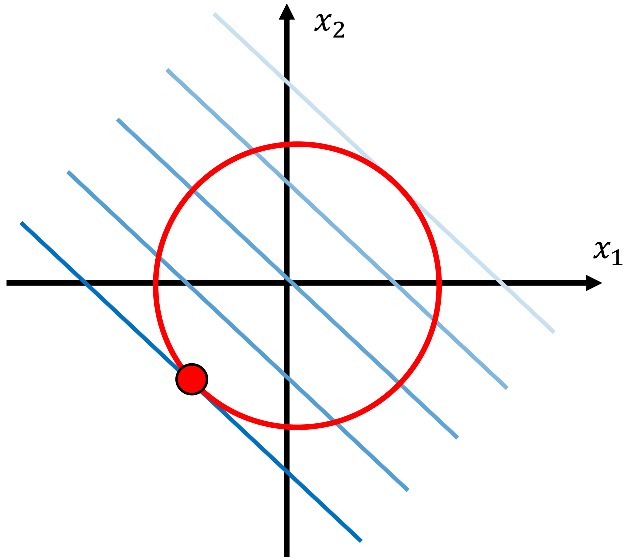
$$H = \begin{bmatrix} L_{x_1x_1} & L_{x_1x_2} \\ L_{x_2x_1} & L_{x_2x_2} \end{bmatrix} = \begin{bmatrix} 2\mu & 0 \\ 0 & 2\mu \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad for\ \mu = 1/2,$$

Positive definite

$$H = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \quad for\ \mu = -1/2,$$

Thus, $\boldsymbol{x}^* = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$



Nonlinear programming: inequality constraints

Contents

- Linear programming
- Nonlinear programming: equality constraints
  - Method of Lagrange Multiplier
- Nonlinear programming: inequality constraints
  - Lagrangian Dual Problem
  - KKT Conditions
  - Example: constrained optimization problem
- Projection-type methods
- Penalty-type methods

Lagrangian Dual Problem

- Consider the following nonlinear problem P, which we call the primal problem:

$$\text{P:} \quad p^* = \text{minimum}_x \quad f(x)$$

$$\text{s.t.} \quad g_i(x) \leq 0, \quad i = 1, \ldots, m$$

$$x \in \Omega$$

Lagrangian function: $L(x, u) := f(x) + u^T g(x)$

Dual function: $L^*(u) := \text{minimum}_x \quad f(x) + u^T g(x)$

$$\text{s.t.} \quad x \in \Omega$$

Dual problem: D: $d^* = \text{maximum}_u \quad L^*(u)$

$$\text{s.t.} \quad u \geq 0$$

Lagrangian Dual Problem

- Theorem: The dual function $L^*(u)$ is a concave function.
- Proof: Let $u_1 \geq 0$ and $u_2 \geq 0$ be two values of the dual variables, and let

$u = \lambda u_1 + (1-\lambda)u_2$, where $\lambda \in [0,1]$. Then

$$\begin{aligned}
L^*(u) &= \min_{x\in P} f(x) + u^T g(x) \\
&= \min_{x\in P} \lambda \left[f(x) + u_1^T g(x) \right] + (1-\lambda) \left[f(x) + u_2^T g(x) \right] \\
&\geq \lambda \left[\min_{x\in P} f(x) + u_1^T g(x) \right] + (1-\lambda) \left[\min_{x\in P} (f(x) + u_2^T g(x) \right] \\
&= \lambda L^*(u_1) + (1-\lambda) L^*(u_2) \quad .
\end{aligned}$$

Therefore we see that $L^*(u)$ is a concave function.

Lagrangian Dual Problem

- Concave functions

Definition 3.12 *A function $f(x)$ is a* concave function *if*

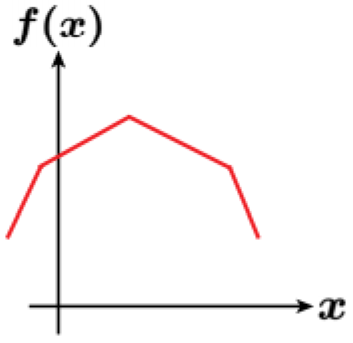
$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y)$$

for all x and y and for all $\lambda \in [0,1]$.

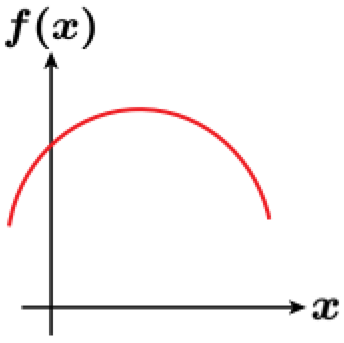
Definition 3.13 *A function $f(x)$ is a* strictly concave function *if*

$$f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda) f(y)$$

for all x and y, $x \neq y$, and for all $\lambda \in (0,1)$.



concave



strictly concave

Lagrangian Dual Problem

KAIST

- Discussion
  - The Lagrangian is affine for all x.
  - The dual is a concave maximization problem.
  - See the example next slide.

Lagrangian function: $L(x, u) := f(x) + u^T g(x)$

Dual function:
$$L^*(u) := \quad \text{minimum}_x \quad f(x) + u^T g(x)$$
$$\text{s.t.} \qquad x \in \Omega$$

Dual problem:
$$\text{D}: \quad \text{maximum}_u \quad L^*(u)$$
$$\text{s.t.} \qquad u \geq 0$$

Lagrangian Dual Problem

- Example
  - Transforms equality constraints into unconstrained problem.

$$f(\mathbf{x}) = 6x_1x_2$$
$$s.t. \quad g(\mathbf{x}) = 3x_1 + 4x_2 = 18.$$

$$L(\mathbf{x},\mathbf{u}) = f(\mathbf{x}) + ug(\mathbf{x}) = 6x_1x_2 + u(3x_1 + 4x_2 - 18)$$

$$\partial L / \partial x_1 = 6x_2 + 3u = 0, \qquad \text{(a)}$$
$$\partial L / \partial x_2 = 6x_1 + 4u = 0, \qquad \text{(b)}$$
$$\partial L / \partial u = 3x_1 + 4x_2 - 18 = 0. \qquad \text{(c)}$$

Lagrangian Dual Problem

KAIST

- Example (continued)
  - First of all, let us check if the dual function $L^*(u)$ is concave.

$$Since \quad x_2 = -\frac{1}{2}u,\ x_1 = -\frac{2}{3}u$$

$$L^*(u) = 6\left(-\frac{2}{3}u\right)\left(-\frac{1}{2}u\right) + u\left(3\left(-\frac{2}{3}u\right) + 4\left(-\frac{1}{2}u\right) - 18\right)$$

$$= 2u^2 - 4u^2 - 18u = -2u^2 - 18u. \quad \longleftarrow \text{Concave!}$$

Lagrangian Dual Problem

- Example (continued)
  - Second, find minimizers.

$$Since \quad x_2 = -\frac{1}{2}u,\ x_1 = -\frac{2}{3}u$$

$$u^* = -4.5 \quad \longleftarrow \quad \text{By (c)}$$

$$thus\ x_1^* = 3,\ x_2^* = \frac{9}{4}.$$

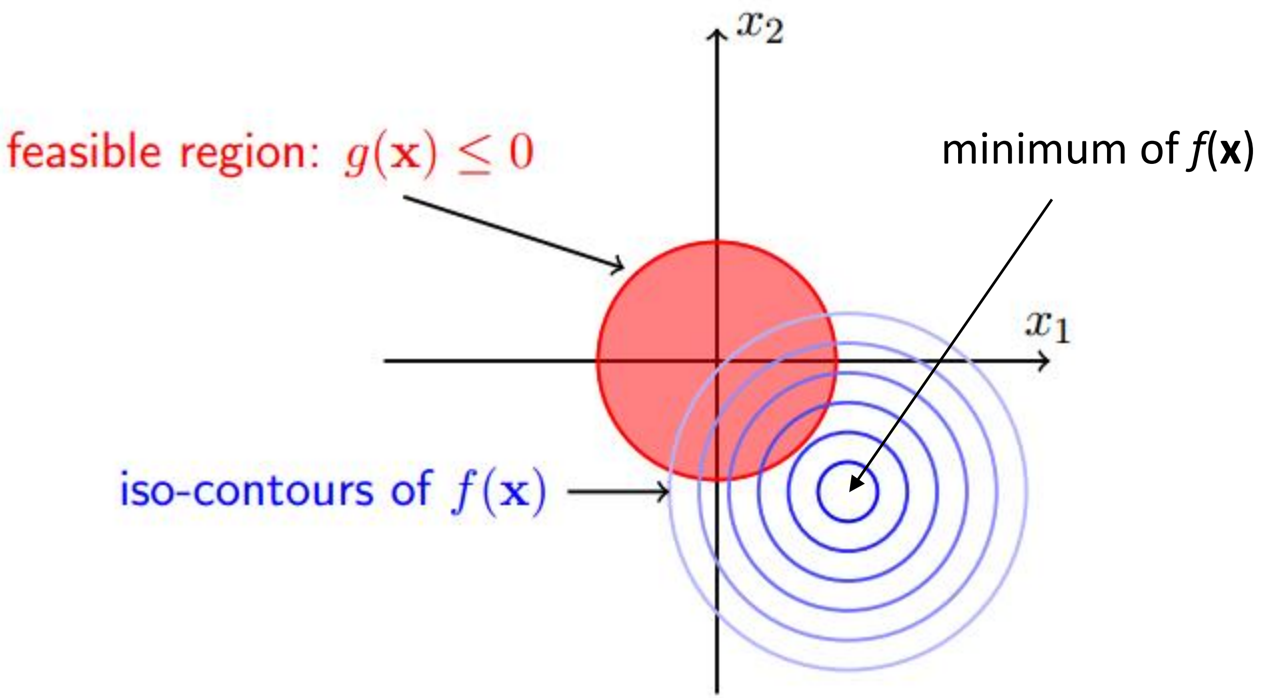
KKT Conditions

- For an optimization problem with differentiable objective and constraint functions, for which strong duality holds, we can derive a set of necessary conditions for optimality, known as the Karuch-Kuhn-Tucker (KKT) conditions.
- If strong duality holds and x^*, (λ^*, ν^*) are a primal- and a dual-optimal solution, relatively, then it holds the KKT conditions.

KKT Conditions

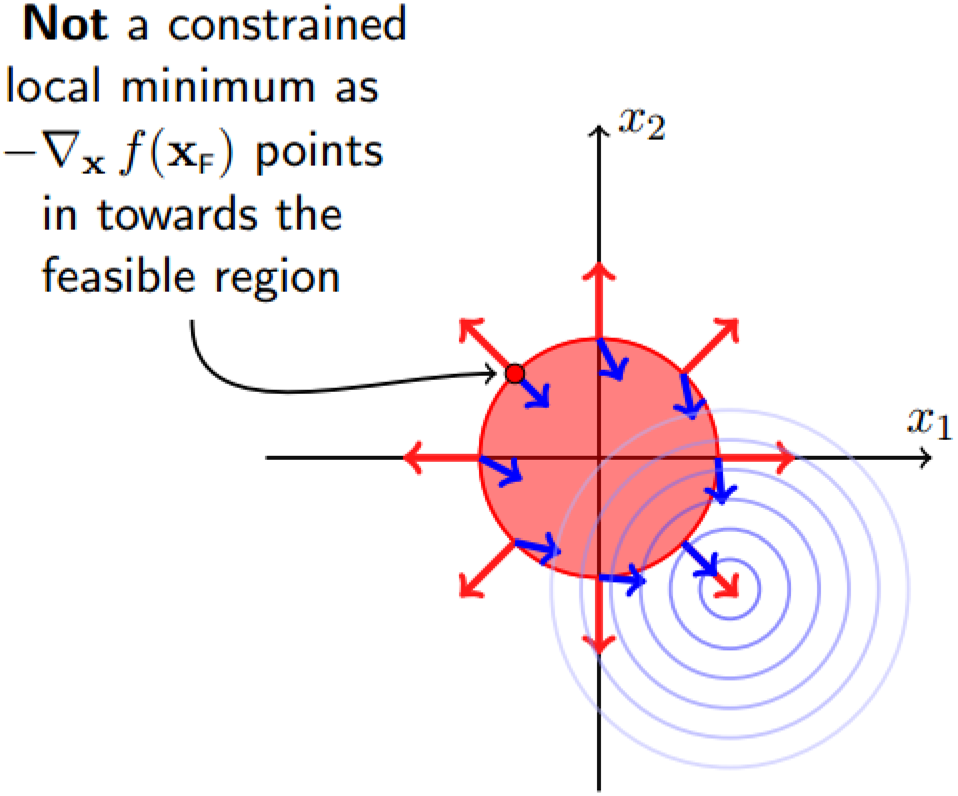
$$\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \le 0$$

where $$f(\mathbf{x}) = (x_1 - 1.1)^2 + (x_2 - 1.1)^2 \textbf{ and } g(\mathbf{x}) = x_1^2 + x_2^2 - 1$$



Equal to an optimization problem with an equality constraint: $g(\mathbf{x})=0$.

KKT Conditions



$\therefore$ Constrained local minimum occurs when $-\nabla_{\mathbf{x}} f(\mathbf{x})$ and $\nabla_{\mathbf{x}} g(\mathbf{x})$ point in the same direction:

$$\boxed{-\nabla_{\mathbf{x}} f(\mathbf{x}) = \lambda \nabla_{\mathbf{x}} g(\mathbf{x}) \quad \textbf{and} \quad \lambda > 0}$$

KKT Conditions

Given

$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \;\text{ subject to }\; g(\mathbf{x}) \leq 0}$$

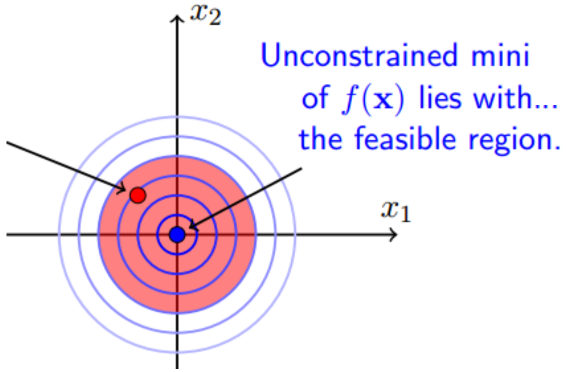
If $\mathbf{x}^*$ corresponds to a constrained local minimum then

Case 1:
Unconstrained local minimum occurs **in** the feasible region.

1. $g(\mathbf{x}^*) < 0$
2. $\nabla_{\mathbf{x}} f(\mathbf{x}^*) = \mathbf{0}$
3. $\nabla_{\mathbf{xx}} f(\mathbf{x}^*)$ is a positive semi-definite matrix.

Case 2:
Unconstrained local minimum lies **outside** the feasible region.

1. $g(\mathbf{x}^*) = 0$
2. $-\nabla_{\mathbf{x}} f(\mathbf{x}^*) = \lambda \nabla_{\mathbf{x}} g(\mathbf{x}^*)$ with $\lambda > 0$
3. $\mathbf{y}^t \nabla_{\mathbf{xx}} L(\mathbf{x}^*)\, \mathbf{y} \geq 0$ for all $\mathbf{y}$ orthogonal to $\nabla_{\mathbf{x}} g(\mathbf{x}^*)$.



KKT Conditions

Given the optimization problem

$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq 0}$$

Define the **Lagrangian** as

$$\boxed{\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda\, g(\mathbf{x})}$$

Then $\mathbf{x}^*$ a local minimum $\iff$ there exists a unique λ^* s.t.

1. $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda^*) = \mathbf{0}$
2. $\lambda^* \geq 0$
3. $\lambda^*\, g(\mathbf{x}^*) = 0$
4. $g(\mathbf{x}^*) \leq 0$
5. Plus positive definite constraints on $\nabla_{\mathbf{xx}}\, \mathcal{L}(\mathbf{x}^*, \lambda^*)$.

These are the **KKT conditions**.

KKT Conditions

- KKT for multiple equality & inequality constraints

Given the constrained optimization problem

$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x})}$$

subject to

$$\boxed{h_i(\mathbf{x}) = 0 \text{ for } i = 1,\ldots,l \;\; \textbf{and} \;\; g_j(\mathbf{x}) \leq 0 \text{ for } j = 1,\ldots,m}$$

Define the **Lagrangian** as

$$\boxed{\mathcal{L}(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\mu}^t\,\mathbf{h}(\mathbf{x}) + \boldsymbol{\lambda}^t\,\mathbf{g}(\mathbf{x})}$$

Then $\mathbf{x}^*$ a local minimum $\iff$ there exists a unique $\boldsymbol{\lambda}^*$ and μ* s.t.

1. $\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*) = \mathbf{0}$
2. $\lambda_j^* \geq 0$ for $j = 1,\ldots,m$
3. $\lambda_j^*\, g_j(\mathbf{x}^*) = 0$ for $j = 1,\ldots,m$
4. $g_j(\mathbf{x}^*) \leq 0$ for $j = 1,\ldots,m$
5. $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$
6. Plus positive definite constraints on $\nabla_{\mathbf{xx}}\,\mathcal{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*)$.

KKT Conditions

- Note that
  - Under differentiability and constraint qualifications, the Karush–Kuhn–Tucker (KKT) conditions provide necessary conditions for a solution in nonlinear programming to be optimal.
  - Under convexity, these conditions are also sufficient.

Nonlinear programming: others

Contents

- Linear programming
- Nonlinear programming: equality constraints
  - Method of Lagrange Multiplier
- Nonlinear programming: inequality constraints
  - Lagrangian Dual Problem
  - KKT Conditions
  - Example: constrained optimization problem
- Projection-type methods
- Penalty-type methods

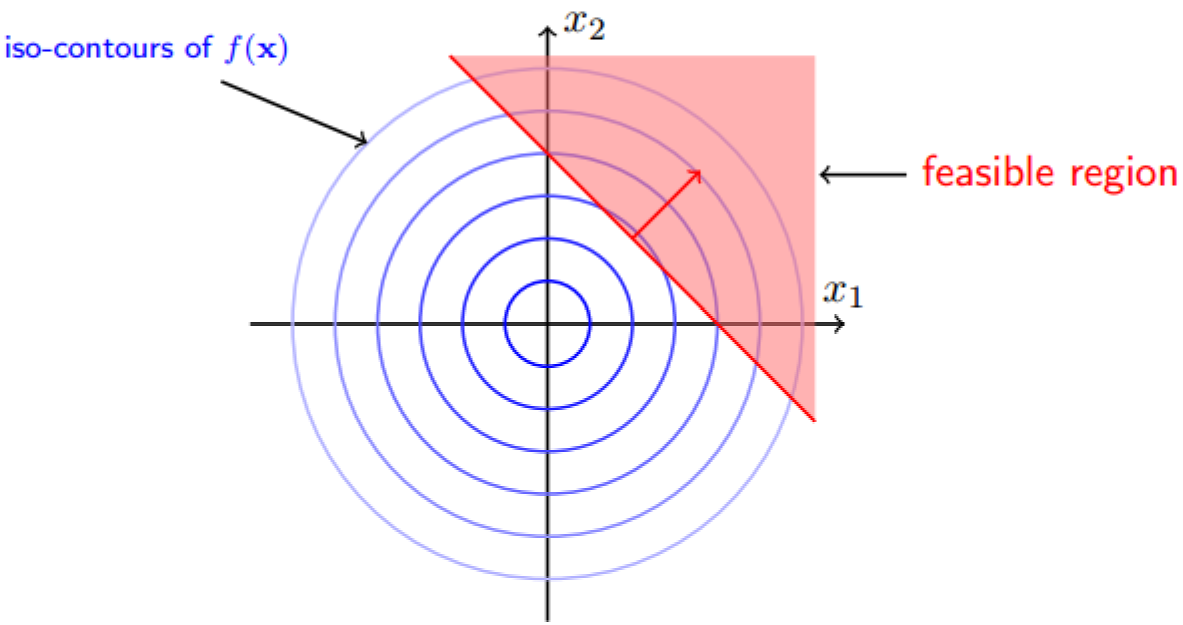
Example: constrained optimization problem

- Solve this constrained optimization problem

$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) = \min_{\mathbf{x}\in\mathbb{R}^2} .4\,(x_1^2 + x_2^2)}$$

subject to

$$\boxed{g(\mathbf{x}) = 2 - x_1 - x_2 \leq 0}$$



Example: constrained optimization problem

- Solution 1: Using KKT conditions

The Lagrangian is

$$\mathcal{L}(\mathbf{x}, \lambda) = .4\,x_1^2 + .4\,x_2^2 + \lambda\,(2 - x_1 - x_2)$$

The KKT conditions say that at an optimum $\lambda^* \geq 0$ and

$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \lambda^*)}{\partial x_1} = .8\,x_1^* - \lambda^* = 0$$
$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \lambda^*)}{\partial x_2} = .8\,x_2^* - \lambda^* = 0$$
$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \lambda^*)}{\partial \lambda} = 2 - x_1^* - x_2^* = 0$$

Example: constrained optimization problem

- Solution 1: Using KKT conditions (continued)

Find $(x_1^*, x_2^*, \lambda^*)$ which fulfill these simultaneous equations. The first two equations imply

$$x_1^* = \frac{5}{4}\lambda^*, \qquad x_2 = \frac{5}{4}\lambda^*$$

Substituting these into the last equation we get

$$8 - 5\lambda^* - 5\lambda^* = 0 \quad \Longrightarrow \quad \lambda^* = \frac{4}{5} \leftarrow \text{greater than 0}$$

and in turn this means

$$x_1^* = \frac{5}{4}\lambda^* = 1, \qquad x_2^* = \frac{5}{4}\lambda^* = 1$$

Example: constrained optimization problem

- Solution 2: Using Lagrangian dual function

Construct the *Lagrangian dual function*

$$q(\lambda) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda) = \min_{\mathbf{x}} \ (f(\mathbf{x}) + \lambda g(\mathbf{x}))$$

Find optimal value of $\mathbf{x}$ wrt $\mathcal{L}(\mathbf{x}, \lambda)$ in terms of the Lagrange multiplier:

$$x_1^* = \frac{5}{4}\lambda, \qquad x_2^* = \frac{5}{4}\lambda$$

Substitute back into the expression of $\mathcal{L}(\mathbf{x}, \lambda)$ to get

$$q(\lambda) = \frac{5}{4}\lambda^2 + \lambda\,(2 - \frac{5}{4}\lambda - \frac{5}{4}\lambda)$$

Find $\lambda \geq 0$ which maximizes $q(\lambda)$. Luckily in this case the global optimum of $q(\lambda)$ corresponds to the constrained optimum

$$\frac{\partial q(\lambda)}{\partial \lambda} = -\frac{5}{2}\lambda + 2 = 0 \quad \Longrightarrow \quad \lambda^* = \frac{4}{5} \quad \Longrightarrow \quad x_1^* = x_2^* = 1$$

Projection-type methods

- Projection-type methods address the problem of optimizing over convex sets.
  - A convex set $\mathcal{C}$ is a set such that

$$\theta x + (1-\theta) y \in \mathcal{C},$$

$$\text{for all } x,\ y \in \mathcal{C} \text{ and } 0 \leq \theta \leq 1.$$

  - Projection-type methods use the projection operator,

$$P_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \frac{1}{2} ||x - y||^2.$$

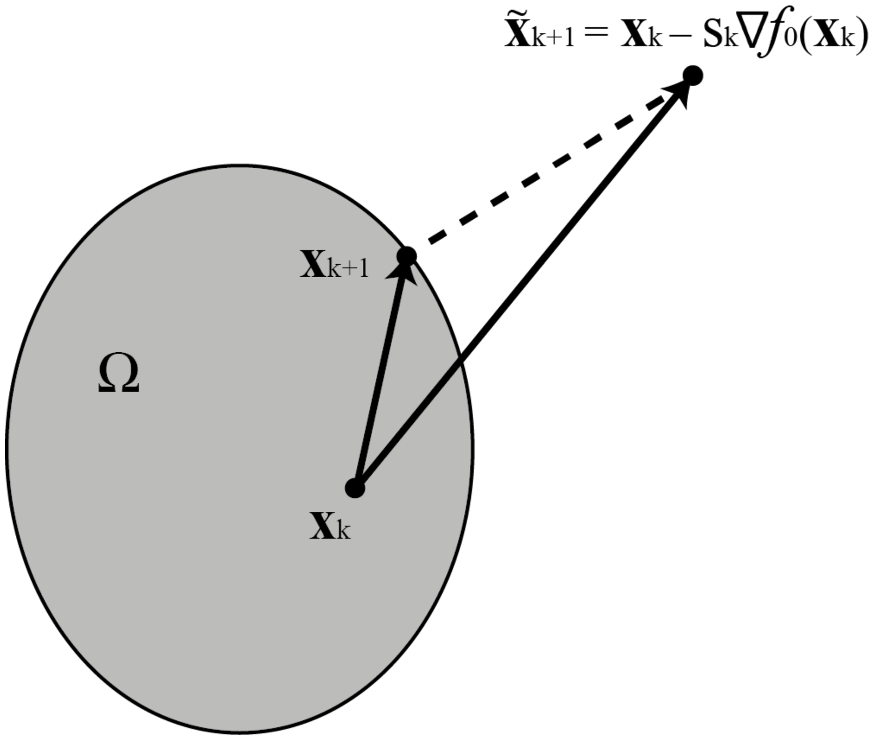
  - For non-negative constraints, this operator is simply

$$x = \max\{0, x\}.$$

Projection-type methods

- The most basic projection-type method is *gradient projection*.

$$x_{k+1} = P_C(x_k - \alpha_k \nabla f(x_k)).$$



Penalty-type methods

- Penalty-type methods re-write as an unconstrained problem, e.g.
  - Penalty method for equality constraints: Re-write

$$\min_{c(x)=0} f(x),$$

as

$$\min_{x} f(x) + \frac{\mu}{2}||c(x)||^2.$$

Penalty-type methods

- Penalty-type methods re-write as an unconstrained problem, e.g.
  - Penalty method for equality constraints: Re-write

$$\min_{c(x)=0} f(x),$$

as

$$\min_{x} f(x) + \frac{\mu}{2}||c(x)||^2.$$

  - Penalty method for inequality constraints: Re-write

$$\min_{c(x)\geq 0} f(x),$$

as

$$\min_{x} f(x) + \frac{\mu}{2}||\max\{0, c(x)\}||^2.$$

These converge to the original problem as $\mu \to \infty$.

Penalty-type methods

Minimize $x_1^2 + x_2^2$

s.t. $x_1 + x_2 - 1 = 0.$

1. By the method of Lagrange multiplier, the optimal solution is (1/2, 1/2) and has the objective value1/2.
2. Now consider the penalty problem

Minimize $x_1^2 + x_2^2 + \mu(x_1 + x_2 - 1)^2$

$$\Rightarrow \quad \begin{aligned} x_1 + \mu(x_1 + x_2 - 1) &= 0 \\ x_2 + \mu(x_1 + x_2 - 1) &= 0. \end{aligned}$$

$$\Rightarrow \quad x_1 = x_2 = \mu/(2\mu + 1).$$

Thus, the optimal solution of the penalty problem can be made arbitrarily close to the solution of the original problem by choosing μ sufficiently large.

Penalty-type methods

KAIST

- Selecting the penalty parameter

$$\min \Phi_k(\mathbf{x}) = f(\mathbf{x}) + \mu_k \sum_{i \in I} g(c_i(\mathbf{x}))$$

  – Start with a relatively small value of μ .
  – Subsequently, solve a sequence of unconstrained problems with monotonically increasing values of μ chosen so that the solution to each new problem is “close” to the previous one.

Table A1. Sequence of Solutions Using the Penalty Method

k	μ	x_1	x_2	g_3
0	—	6.00	7.00	6.00
1	0.5	4.50	5.50	3.00
2	1	4.00	5.00	2.00
3	2	3.60	4.60	1.20
4	4	3.33	4.33	0.66
5	8	3.18	4.18	0.35
6	16	3.09	4.09	0.18
7	32	3.05	4.05	0.09
8	64	3.02	4.02	0.04

References

- An introduction to optimization, by Chong, 3rd Ed., Wiley
- Optimization for computer vision, by Treiber, Springer
- Nonlinear programming, by Bazaraa et al., 3rd Ed., Wiley
- www.wikipedia.com